UDK: 551.577.22
Stručni rad

STATISTICAL ANALYSIS OF WET AND DRY SPELLS IN CROATIA BY THE BINARY DARMA (1,1) MODEL

Statistička analiza sušnih i kišnih razdoblja u Hrvatskoj primjenom modela DARMA (1,1)

CINDRIĆ KSENIJA

Meteorological and Hydrological Service
Grič 3, 10000 Zagreb, Croatia
cindric@cirus.dhz.hr

Prihvaćeno 9. listopada 2006, u konačnom obliku 25. siječnja 2007.

Abstract: In this work the Discrete Autoregressive Moving Average Model (DARMA(1,1)) has been used to describe the wet-dry day sequences in Croatia. The dryness or wetness of a time period (e.g. a month) can be determined using daily precipitation data. Nevertheless, in some cases these estimates may give a wrong impression, and therefore monitoring the number of wet and dry days is important as well. The First-order Markov Chain (DAR(1)) has also been used and compared to DARMA(1,1) results. The daily precipitation data from the Split-Marjan, Rijeka, Zagreb-Grič and Osijek meteorological stations during the 1948–2000 period have been used. The importance of the autocorrelation coefficient as a measure of persistence has been emphasized.

Key words: wet and dry spells, DARMA (1,1) model, DAR (1) model, coefficient of persistency

Sažetak: U radu su analizirani sušni i kišni sljedovi u Hrvatskoj primjenom modela autoregresije i kliznog srednjaka (DARMA(1,1)). Sušnost ili kišnost nekog razdoblja može se ocijeniti na osnovi količina oborine. Međutim, takve procjene u pojedinim slučajevima mogu dati iskrivljenu sliku ako je npr. velika količina oborine pala za malen broj dana, a nakon takva kišnog razdoblja slijedilo je dugačko sušno razdoblje. Zbog toga je monitoring broja dana s oborinom jednako važan kao i praćenje količina oborine te služi za procjenu sušnosti ili kišnosti pojedinog mjeseca. Primijenjen je i jednostavan model Markovljevih lanaca prvog reda (DAR(1)) te su rezultati uspoređeni s rezultatima modela DARMA(1,1). Analiza je napravljena na osnovi podataka dnevne količine oborine za postaje Split-Marjan, Rijeka, Zagreb-Grič i Osijek iz razdoblja 1948–2000. U radu je naglašena i važnost koeficijenta autokorelacije kao mjere perzistencije u sljedovima dana s oborinom.

Ključne riječi: sušna i kišna razdoblja, model DARMA (1,1), model DAR(1), koeficijent perzistencije

1. INTRODUCTION

Precipitation is one of the most important climatic elements having direct impact on human life. Extreme precipitation events, causing floods and droughts, are of vital interest. The great spatial and temporal variability of precipitation amounts make the analysis of the precipitation regime more complicated. Also, there is a large number of statistical values that describe this variability. The dryness or wetness of a particular month can be determined using daily precipitation amount data. However, in some cases these estimates may give a wrong impression. For example, if there was a large amount of precipitation only in a few days‘ period, followed by a very long dry spell, one could conclude that the month was very wet. Hence, besides precipitation amounts, the monitoring of the number of wet and dry spells is also very important in estimating the dryness or wetness of a particular month. Long wet and dry spells are of more interest than short ones. Relying only on empirical frequency distributions may create a completely wrong idea about the probability of very long sequences. Various statistical models give more reliable estimates. In this work, the Discrete Autoregressive Moving Average Model (DARMA(1,1)) and the Discrete Autoregressive (DAR(1)) Model have been used to analyse the distribution of wet and dry spells frequencies and their results have then been compared. The DAR(1) model is known as the First-order Markov Chain. It was first introduced in meteorology by Gabriel and Neumann (1962). In this model, the probability of a wet or a dry day depends on the situation on the previous day only. Many authors have shown that the properties of wet and dry sequences in climatic areas with very long dry periods can not be described properly by this simple model. For general application, the Markov chains of higher order could be used, but they have the disadvantage of an increased number of parameters. Hence, the more complex DARMA(1,1) model has been used. In climatic analysis, this model was first introduced by Buishand (1978). The addition of one more parameter provides a better fit of this model to empirical values than DAR(1) does.

This work deals with the daily precipitation data measured at four meteorological stations in Croatia, during the 1948–2000 period: Split-Marjan (φ = 43°31', λ = 16°26', h = 122 m), on the mid-Adriatic coast of Croatia, Rijeka (φ = 45°20', λ = 14°27', h = 120 m), on the northern Adriatic coast, Zagreb-Grič (φ = 45°49', λ = 15°59', h = 157 m), in Croatian hinterland and Osijek (φ = 45°30', λ = 18°34', h = 89 m), in the eastern part of Croatia. These stations are representative of the different types of climate existing in Croatia. The importance of the autocorrelation coefficient (acc) in describing precipitation amount variability is emphasized. The annual courses of the acc for the mentioned stations are presented.

2. THE BINARY DARMA(1,1) MODEL

2.1. Some properties of the DAR(1) and DARMA(1,1) models

The DAR(1) and DARMA(1,1) models are special cases of the more general class of DARMA(p,q) models. The main property of the DARMA models is that their marginal distribution and correlation structure are specified separately.

A DARMA sequence $\{X_n\}$ is formed by a probabilistic linear combination of a sequence $\{Y_n\}$ of independent and identically distributed random variables in such a way that its marginal distribution is given by: $\pi(k) = P(Xn = k)$, k = 0,1,... In a binary discrete sequence, these variables take the value one with $\pi(1)$ probability and the value zero with $\pi(0) = 1 - \pi(1)$ probability.

The random variable A_n in the DAR(1) model (First-order Markov Chain) is defined as follows:

$A_n = A_{n-1}$ with ρ probability, $A_n = Y_n$ with 1-ρ probability. The autocorrelation function (acf) for this process is given by:

$$\rho_k = \rho_1^k, \quad k \geq 1 \tag{1}$$

In the DARMA(1,1) process, X_n takes the value Y_n with β probability and takes the autoregressive component A_{n-1} with $1-\beta$ probability. This model is completely determined by three parameters: $\pi(1)$, ρ and β. The autocorrelation function for this process is given by:

$$\rho_k = c\rho^{k-1}, \ k \geq 1 \tag{2}$$

where c is the first-order autocorrelation coefficient (acc). This indicates that c is a valuable value which determines acf. The acf of the daily precipitation sequence is a measure of persistence, which is one of the most important statistical properties when considering dry and wet spell length. Then, c can also be taken as a model parameter rather than β.

It should be noted that the DARMA(1,1) process $\{X_n\}$ is not Markovian but $\{A_n, X_n\}$ is a bivariate first-order Markov chain.

2.2. Estimation of parameters

While the DAR(1) model is completely determined by the two parameters $\pi(1)$ and c, the DARMA(1,1) process has an additional parameter, ρ, which gives the amount of the acf decrease. If one has the empirical distribution of the run lengths of dry and wet spells, the estimates of the marginal distribution $\pi(1)$ and c can be based on the mean run lengths, μ_0 and μ_1 as follows (Buishand, 1978):

$$\pi(1) = \frac{\mu_1}{\mu_1 + \mu_0} \tag{3}$$

$$c = 1 - \frac{1}{\mu_0} - \frac{1}{\mu_1} \tag{4}$$

There are several ways of estimating parameter ρ. In this work, the parameter estimation is based on the distribution of run lengths (Buishand, 1978). The difference between the first and the second serial correlation coefficients r_1 and r_2 can be expressed as a monotonically decreasing function of the sum of relative frequencies of runs of zeros with length 1, $f_0(1)$ and runs of ones with length 1, $f_1(1)$, as follows:

$$r_1 - r_2 = \frac{\mu_1 + \mu_0}{\mu_1 \mu_0}\left[1 - f_0(1) - f_1(1)\right] \tag{5}$$

Then, parameter ρ is given a

$$\rho = \frac{r_2}{r_1} \tag{6}$$

An estimate of β is then obtained from the estimates of c and ρ using:

$$c = (1-\beta)\ (\rho+\beta-2\rho\beta) \tag{7}$$

These expressions are briefly derived in Buishand (1978).

2.3. The distribution of run lengths

Dry and wet run lengths are of great importance for many practical purposes (agricultural systems, water supply and irrigation in water systems). The probabilities of occurrence of different daily precipitation sequence lengths are also of great interest.

A run length of zeroes is defined as a sequence of zeroes on each side bounded by a one, and a run length of ones is defined in an analogous manner. If T_0 is the run length of zeroes then the probability distribution of relative frequencies for the First-order Markov Chain can be obtained using conditional probabilities as follows:

$$P(T_0 = n) = p_{00}^{\ n-1} \tag{8}$$

where $p_{00} = \rho + (1-\rho)\pi(0)$

Similarly, for the run lengths of ones, one can obtain

$$P(T_1 = n) = p_{11}^{\ n-1} \tag{9}$$

where $p_{11} = \rho + (1-\rho)\pi(1)$ .The theoretical probability distribution functions of run lengths for the binary DARMA models are given in detail in Chang et al. (1984a, 1984b). The transition probabilities are 2x2 matrices and can be obtained by model parameters which simplify their calculation. Nevertheless, these calculations are more complicated in the DARMA(1,1) than in the DAR(1) model because of a larger number of parameters.

3. RESULTS

The 53-year daily precipitation data (1948–2000) at four meteorological stations in Croatia have been analysed. Wet or dry days have been classified according to whether there was a recorded 1 mm of precipitation during the day. Then, these data were transformed into a binary discrete time series in the following manner:

$$X_n = \begin{cases} 1, & \text{if } R_n \geq 1 \text{ mm} \\ 0, & \text{if } R_n < 1 \text{ mm,} \end{cases}$$

where R_n is the observed daily precipitation amount.

When analysing the wet and dry spells in a particular month, all spells beginning in that month were taken into account even if they continued into the next month. If, for example, July was the month studied, then a dry or a wet spell which began in July and ended in August was taken into account in the July study and discarded from August.

According to these runs, the mean run lengths of dry (μ_0) and wet (μ_1) spells were determined. The probability $\pi(1)$ and the first-order autocorrelation coefficient c ($=\rho_1$) were then obtained by Equations (3) and (4). The parameter ρ was estimated from the runs of zeroes of one $f_0(1)$ length and the runs of ones of one $f_1(1)$ length, using Equations (5) and (6). The third parameter, β was obtained using Equation (7). An example of the estimation of parameters and of the empirical and theoretical frequencies of dry spells for the Rijeka station in January is given in Table 1. The final estimates of the mean run lengths of dry (μ_0) and wet (μ_1) spells, parameters $\pi(1)$, c, ρ and β, the longest dry (d_0) and wet (d_1) spells for all months and every station are given in

Table 1. The empirical and theoretical frequencies of dry spells and an example of the estimation of parameters $\pi(1)$, c, ρ and β using equations (3) to (7), in January, for the Rijeka station.

Tablica 1. Empiričke i teoretske čestine sušnih razdoblja i primjer izračunavanja parametara $\pi(1)$, c, ρ i β dobivenih iz jednadžbi (3) do (7) za siječanj na postaji Rijeka.

DRY SPELLS					WET SPELLS			
n	Emp	n*Emp	DARMA(1,1)	DAR(1)	Emp	n*Emp	DARMA(1,1)	DAR(1)
1	56	56	49.0	36.6	88	88	95.9	89.4
2	25	50	26.8	30.0	55	110	45.4	49.6
3	24	72	21.2	24.6	22	66	26.0	27.6
4	9	36	17.6	20.2	16	64	15.0	15.3
5	17	85	14.6	16.5	8	40	8.7	8.5
6	9	54	12.2	13.5	4	24	5.0	4.7
7	13	91	10.2	11.1	6	42	2.9	2.6
8	7	56	8.5	9.1	1	8	1.7	1.5
9	6	54	7.1	7.5	0	0	1.0	0.8
10	4	40	5.9	6.1	1	10	0.6	0.4
11	7	77	5.0	5.0				
12	3	36	4.1	4.1				
13	3	39	3.4	3.4				
14	1	14	2.9	2.8				
15	2	30	2.4	2.3				
16	4	64	2.0	1.9				
17	2	34	1.7	1.5				
18	4	72	1.4	1.2				
19	0	0	1.2	1.0				
20	0	0	1.0	0.8				
21	3	63	0.8	0.7				
22	1	22	0.7	0.6				
23	1	23	0.6	0.5				
24	0	0	0.5	0.4				
25	0	0	0.4	0.3				
26	0	0	0.3	0.3				
27	0	0	0.3	0.2				
28	0	0	0.2	0.2				
29	2	58	0.2	0.1				
Σ	203	1126			201	452		
	μ_0	5.547			**μ_1**	2.249		
	$f_0(1)$	0.276			**$f_1(1)$**	0.438		
	$\pi(0)$	0.712	Eq. (1) - (3)		**$\pi(1)$**	0.288	Eq. (3)	
c	**0.375**	Eq.(4)						
r_1-r_2	**0.179**	Eq.(5)						
ρ	**0.523**	Eq.(6)						
β	**0.266**	Eq.(7)						

Table 2. The mean run lengths of dry and wet spells have very close values at all stations throughout the year (four to six days). The exception is Split, where the dry spell lengths are considerably greater from July to September with a noticeable peak in August (12 days). In some cases (e.g. April and September in Split) parameter β can have a complex value.

In Figure 1, the annual courses of the first-order acc (c) for four stations, during the period

Table 2. The mean run lengths of dry (μ_0) and wet (μ_1) spells, estimated parameters $\pi(1)$, c, ρ and β for the binary DARMA(1,1) model and the longest dry (d_0) and wet (d_1) spells during the period 1948–2000.

Tablica 2. Srednja duljina trajanja sušnih (μ_0) i kišnih razdoblja (μ_1), parametri $\pi(1)$, c, ρ i β binarnog modela DARMA(1,1) i najdulje sušno (d_0) i kišno (d_1) razdoblje u periodu 1948–2000.

	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII
						OSIJEK						
μ_0	5.522	4.766	5.277	4.448	4.367	3.996	5.196	5.350	6.054	6.841	4.559	4.391
μ_1	1.671	1.704	1.691	1.835	1.713	1.813	1.550	1.570	1.530	1.779	1.897	1.876
d_0	37	24	31	35	36	28	32	30	39	34	51	64
d_1	8	7	8	8	7	7	6	7	6	9	7	9
$\pi(1)$	0.232	0.263	0.243	0.292	0.282	0.312	0.230	0.227	0.202	0.206	0.294	0.299
c	0.220	0.203	0.219	0.230	0.187	0.198	0.162	0.176	0.181	0.292	0.253	0.239
ρ	0.327	0.554	0.219	0.223	0.323	0.189	0.489	-0.035	0.275	0.381	0.385	0.322
β	0.583	0.568	0.598	0.022	0.664	0.021	0.677	0.253	0.691	0.382	0.491	0.532
						ZAGREB-GRIČ						
μ_0	6.022	5.363	5.060	4.589	4.318	3.849	4.493	4.617	5.544	5.916	4.949	5.092
μ_1	1.735	1.721	1.742	1.915	1.912	1.905	1.652	1.656	1.775	1.890	2.000	2.018
d_0	44	35	39	34	23	21	21	31	35	30	36	30
d_1	9	7	6	8	7	7	8	9	7	8	7	8
$\pi(1)$	0.224	0.243	0.256	0.294	0.307	0.331	0.269	0.264	0.243	0.242	0.288	0.284
c	0.258	0.233	0.228	0.260	0.245	0.215	0.172	0.180	0.256	0.302	0.298	0.308
ρ	0.317	0.367	0.319	0.385	0.147	0.207	0.065	0.230	0.291	0.329	0.270	0.357
β	0.475	0.522	0.563	0.474	0.264	0.022	0.160	0.706	0.478	0.301	-	0.308
						RIJEKA						
μ_0	5.547	6.200	5.764	5.116	5.084	4.214	5.996	5.412	5.791	5.462	5.127	6.126
μ_1	2.249	2.276	2.116	2.312	1.987	1.783	1.458	1.689	2.110	2.392	2.552	2.380
d_0	29	38	35	31	33	30	32	31	45	38	34	63
d_1	10	10	10	15	10	8	6	7	8	13	10	8
$\pi(1)$	0.288	0.264	0.269	0.311	0.281	0.301	0.196	0.238	0.267	0.305	0.332	0.280
c	0.375	0.399	0.354	0.372	0.300	0.211	0.147	0.223	0.353	0.399	0.413	0.417
ρ	0.523	0.579	0.426	0.418	0.304	0.269	0.211	0.273	0.428	0.512	0.482	0.532
β	0.266	0.270	0.231	0.163	0.263	0.021	0.777	0.587	0.235	0.213	0.153	0.198
						SPLIT - MARJAN						
μ_0	5.435	5.653	5.863	5.871	6.014	7.430	11.866	8.541	6.972	5.557	4.624	4.983
μ_1	2.028	2.020	1.976	1.957	1.599	1.510	1.420	1.458	1.695	1.848	2.292	2.299
d_0	35	53	37	27	28	84	51	48	45	30	42	47
d_1	7	9	7	7	5	5	5	5	9	8	11	9
$\pi(1)$	0.272	0.263	0.252	0.250	0.210	0.169	0.107	0.146	0.196	0.250	0.331	0.316
c	0.323	0.328	0.323	0.319	0.208	0.203	0.211	0.197	0.267	0.279	0.347	0.364
ρ	0.357	0.576	0.426	0.260	0.171	0.340	0.195	0.263	0.185	0.431	0.441	0.466
β	0.243	0.369	0.316	-	0.087	0.624	0.042	0.657	-	0.431	0.264	0.245

1948–2000, are shown. It is very useful to analyse acc since it describes acf, which is a measure of persistency. A similar analysis was made by Juras (1997) who calculated the acc for ten stations in Croatia during a shorter period (1951–1980). It can be seen that at all stations the value of parameter c is greater in the cold part of the year (October to March) than in the warm period (April to September). This can be explained by the fact that in the cold period macro synoptic processes with long wavelengths prevail. They are of greater persistency in the cold part of the year compared to the warm part. In the summer months, mesoscale convective processes prevail, having smaller persistency. Nevertheless, it is obvious from Figure 1 that at the coastal stations (Rijeka and Split) persistency is considerably greater from January to July than at the continental stations (Zagreb and Osijek). Also, we can see that the annual course of the acc in Rijeka and Split has a well defined wave-shape with a noticeable minimum in the summer months, which is not so pronounced in Zagreb and Osijek.

Juras and Jurčec (1976) investigated the acc at the Zagreb-Grič Observatory and pointed out that c was one of the crucial parameters in describing the temporal variability of precipitation amounts.

Table 2 shows that c and ρ have very close values in some months and, therfore, the differences between the DAR(1) and DARMA(1,1) model estimates of run lengths will be small. One month of each season of the year (January, April, July and October) have been chosen to show the difference between these two models and the advantages of the more complicated DARMA(1,1) model for long spells. The empirical and theoretical cumulative frequencies of dry spells at all stations are shown in Figures 2 to 5. Split-Marjan is an exception because the parameter β for April is a complex

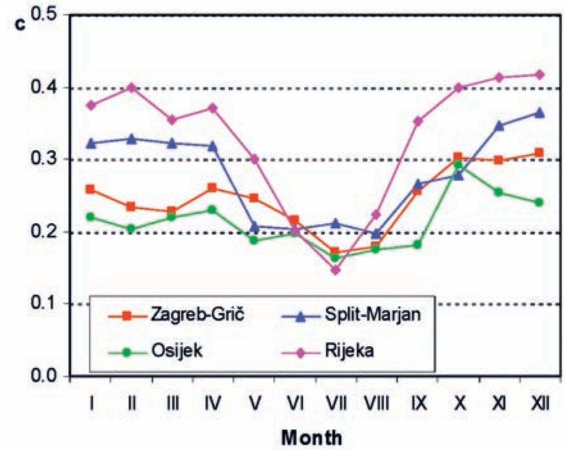

Figure 1. The annual course of the first-order autocorrelation coefficient (c) during the period 1948–2000.

Slika 1. Godišnji hod koeficijenta autokorelacije prvog reda (c) u periodu 1948–2000.

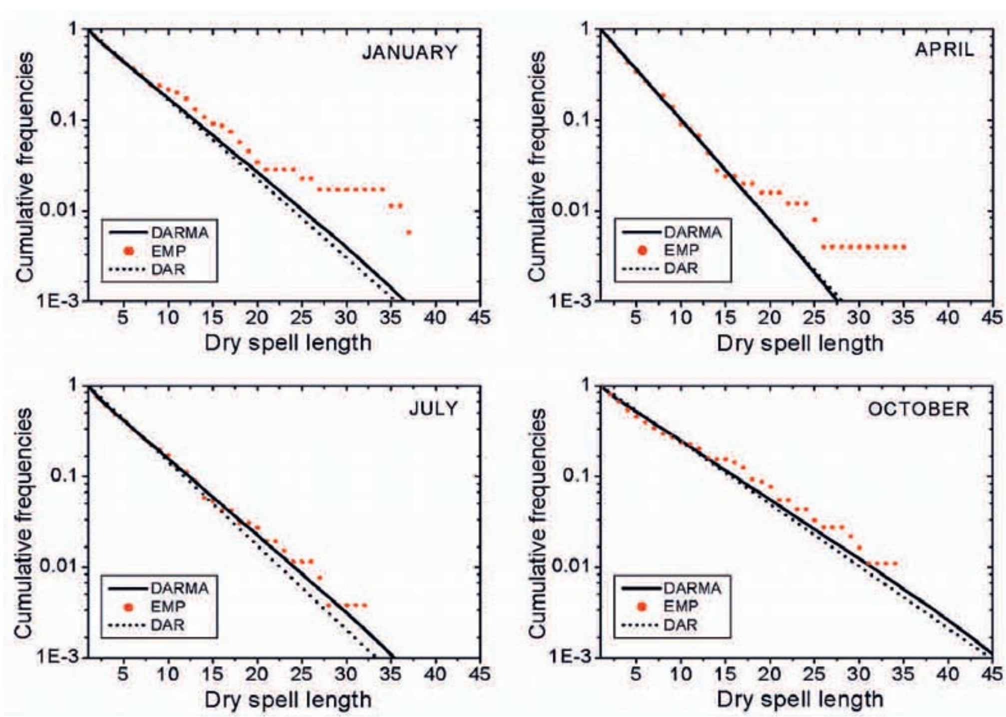

Figure 2. The empirical and theoretical cumulative frequencies of dry spells in January, April, July and October at the Osijek station, 1948–2000.

Slika 2. Empiričke i teoretske kumulativne čestine sušnih razdoblja za siječanj, travanj, srpanj i listopad na postaji Osijek, 1948–2000.

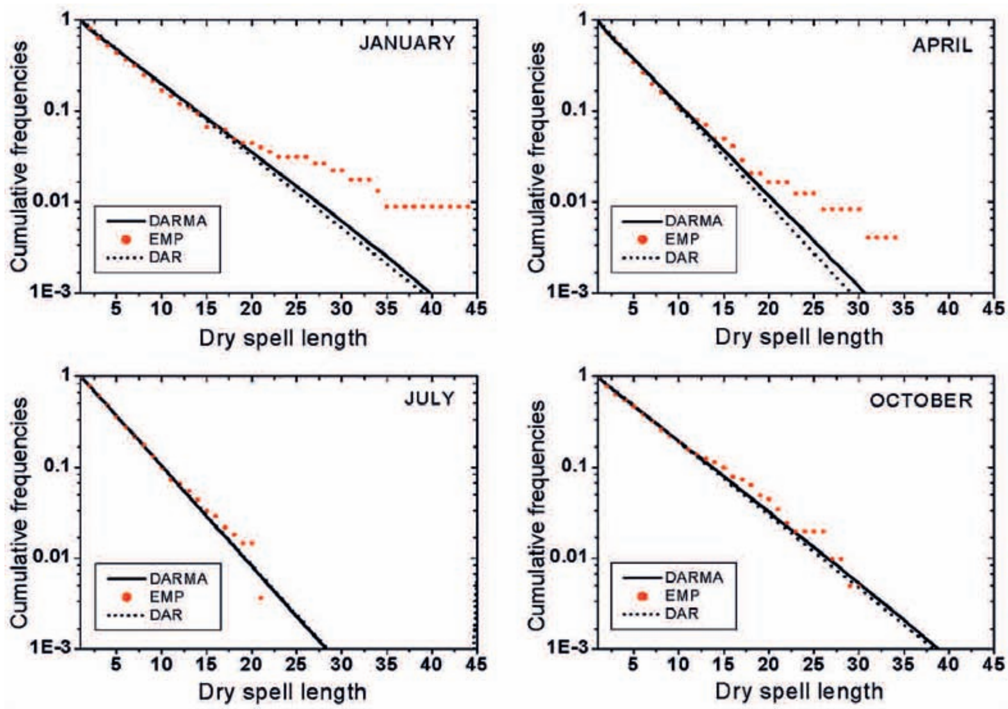

Figure 3. The empirical and theoretical cumulative frequencies of dry spells in January, April, July and October at the Zagreb-Grič station, 1948–2000.

Slika 3. Empiričke i teoretske kumulativne čestine sušnih razdoblja za siječanj, travanj, srpanj i listopad na postaji Zagreb–Grič, 1948–2000.

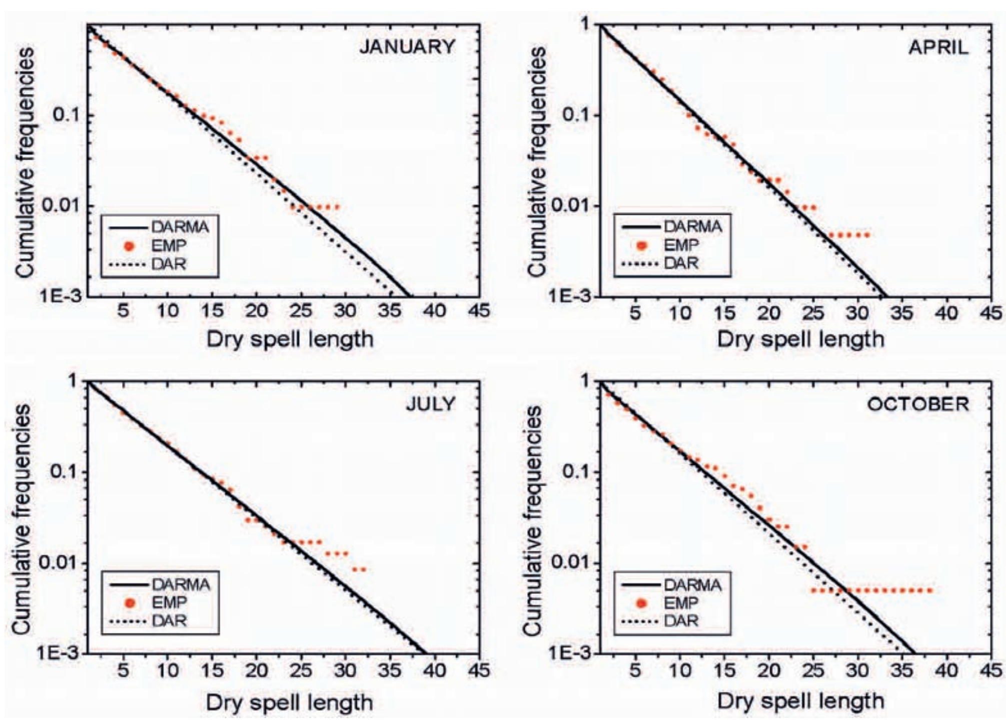

Figure 4. The empirical and theoretical cumulative frequencies of dry spells in January, April, July and October at the Rijeka station, 1948–2000.

Slika 4. Empiričke i teoretske kumulativne čestine sušnih razdoblja za siječanj, travanj, srpanj i listopad na postaji Rijeka, 1948–2000.

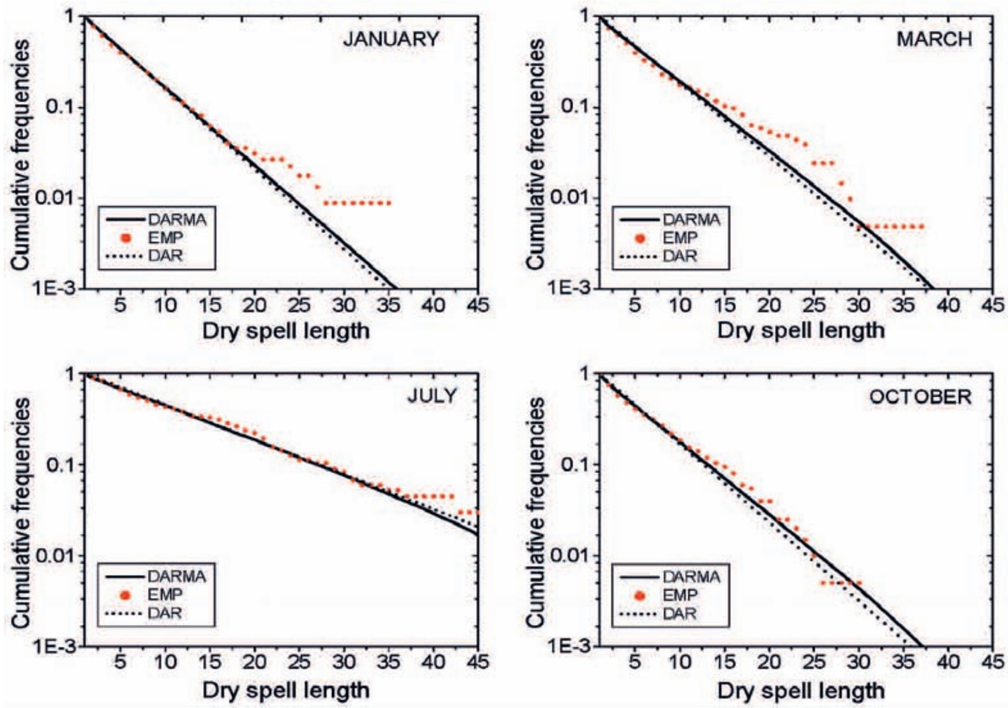

Figure 5. The empirical and theoretical cumulative frequencies of dry spells in January, March, July and October at the Split-Marjan station, 1948–2000.

Slika 5. Empiričke i teoretske kumulativne čestine sušnih razdoblja za siječanj, ožujak, srpanj i listopad na postaji Split-Marjan, 1948–2000.

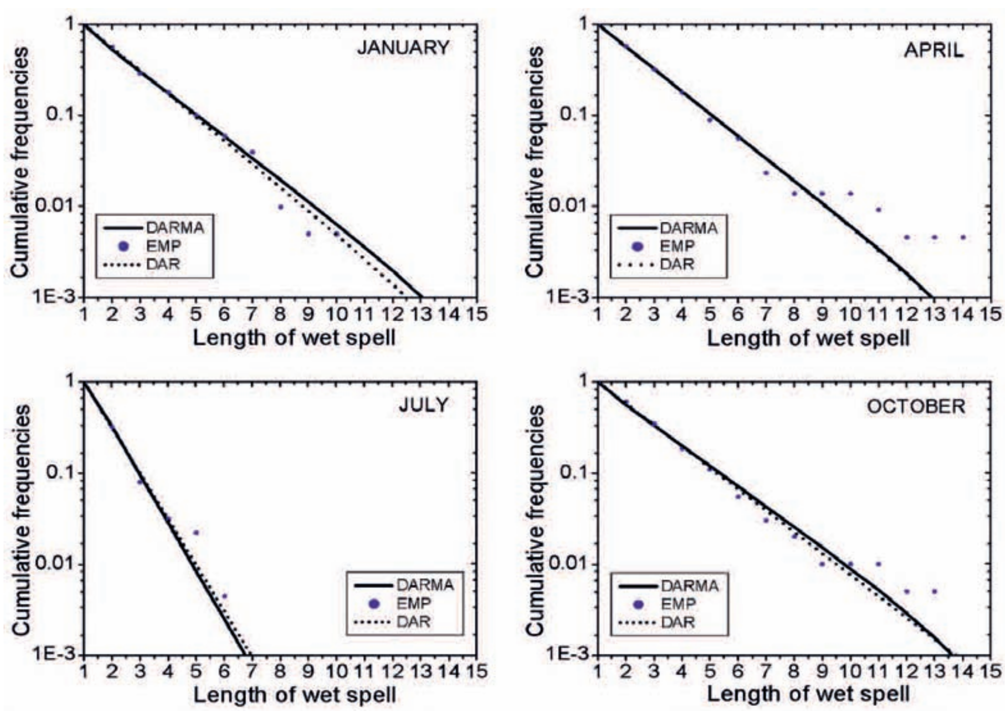

Figure 6. The empirical and theoretical cumulative frequencies of wet spells in January, April, July and October at the Rijeka station, 1948-2000.

Slika 6. Empiričke i teoretske kumulativne čestine kišnih razdoblja za siječanj, travanj, srpanj i listopad na postaji Rijeka, 1948–2000.

number, so March was chosen to show the difference between the two models in Figure 5. In Figure 6, the empirical and theoretical cumulative frequencies of wet spells at Rijeka station are given. It is obvious that, for the Croatian stations, the DARMA(1,1) model gives better estimates for long-term dry spells (more than 10 days), which are more important than short ones. Still, as can be seen from Figures 2 to 6, in most cases the empirical frequencies of very long spells are greater than the theoretical ones. But it can not be said that this discrepancy can be blamed on the models only, because such very long sequences occur rarely, maybe once in 100 years, so that the possibility of such events is very small.

4. CONCLUSIONS

The study of extreme precipitation events such as floods and droughts requires the analysis of wet and dry day sequences. The empirical frequency distribution does not provide a good insight in the probabilities of very long sequences of dry and wet spells, which are more important for practical purposes than the short ones. Statistical models give more reliable estimates. In this work, the dry and wet spells have been studied using the DAR(1) (First-order Markov Chain) and DARMA(1,1) models. The daily precipitation data at four meteorological stations, being representative of different climate regimes in Croatia, have been analysed. Wet or dry days have been classified according to daily precipitation amounts of 1.0 mm.

The results show that both models underestimate very long dry spell runs. However, the DARMA(1,1) model provides a better fit to the empirical distribution both of short (one day) and long (more than 10 days) dry spells. A potential limitation when using the DARMA (1,1) model is the possibility of the appearance of an imaginary parameter β. In this case, other methods of parameter estimation should be used rather than those used in this work (Eq. 5).

Due to the large temporal variability of precipitation amounts, analysis of the precipitation regime is very complex. It is emphasized in this work that the autocorrelation coefficient, as a measure of persistence of wet and dry sequences, is one of the most important parameters of the precipitation regime. The annual course of this parameter is given for all stations. The autocorrelation coefficient is greater in the cold part of the year (October to March) than in the warm period (April to September), which is explained by the atmospheric processes of different persistence prevailing in these parts of the year. In the first half of the year, persistency is greater on the coastal regions of Croatia than in the continental ones.

Acknowledgement: *The author wishes to thank D.Sc. Josip Juras for many valuable suggestions.*

REFERENCES

Buishand, R.A., 1978: The binary DARMA (1,1) process as a model for wet and dry sequence. Dept. of Math., Agricultural University, Wageningen, 49 pp.

Chang, T. J., M. L. Kavvas and J. W. Delleur, 1984a: Daily precipitation modelling by discrete autoregressive moving average processes. *Water Resour. Res.*, **20**, 565–580.

Chang, T. J., M. L. Kavvas and J. W. Delleur, 1984b: Modeling of sequences of wet and dry days by binary discrete autoregressive moving average processes. *J. Clim. Appl. Meteor.*, **23**, 1367–1378.

Gabriel, K. R. and J. Neumann, 1962: A Markov chain model for daily rainfall occurrence at Tel Aviv. *Quart. J. Roy. Meteor. Soc.*, **88**, 90–95.

Juras, J. and V. Jurčec, 1976: The statistical analysis of dry and wet spells by the application of Markov chain probability model. *Papers/Rasprave i prikazi*, **13,** RHMZ Hrvatske, 59–98 (*in Croatian*).

Juras, J., 1989: On modelling binary meteorological sequences with special emphasis on frequencies of warm and cold spells. *Papers/Rasprave*, **24**, 29–37 (*in Croatian*).

Juras, J. 1997: Annual variation of persistency in the sequences of days with precipitation. *The Adriatic meteorology/Jadranska meteorologija*, **XLII**, 9–14.